Prognosis of Rail Transport Demand in The Slovak Republic
Research of Input Parameters

Jozef Danis
Faculty of Operation and Economics of Transport and Communications, University of Zilina
Zilina, Slovakia
jozef.danis@fpedas.uniza.sk

Anna Dolinayova
Faculty of Operation and Economics of Transport and Communications, University of Zilina
Zilina, Slovakia
anna.dolinayova@fpedas.uniza.sk

Juraj Camaj
Faculty of Operation and Economics of Transport and Communications, University of Zilina
Zilina, Slovakia
juraj.camaj@fpedas.uniza.sk

Abstract— Nowadays modern management cannot do without mathematical and statistical tools. Managers need a variety of market information, but they must have inside information about the company that will make their decisions relevant. This paper describes the possibility of creating a regression model of transport requirements in railway transport, depending on other economic indicators. It deals with the problem of multicollinearity among input parameters, which are needed for the creation of a prognosis.

Keywords— regression model, multicollinearity, demand management, railway transport, transport requirements, economic indicators

I. INTRODUCTION

Globalization of economics requires rail company managers to make many decisions. A lot of information on the future development of monitored parameters is needed for each decision. Managers have to make estimates about transport performances at the creation of operation plans because these influence sales, individual cost items, utilization of rolling stocks and railway infrastructure and so on. They can estimate dominant trends and seasonal effects or identify interdependencies between variables and the estimates of necessary data or correct their own human judgment from the amount of information available to them for a certain period of time.

II. 1. INPUT DATA FOR THE CREATION OF A REGRESSIVE MODEL FOR RAILWAY TRANSPORT DEMAND

Demand is an abstract concept reflecting what individuals would like to consume under different scenarios. It is generally considered that the quantity demanded of a commodity is influenced by its price, the price of other goods, tastes and the level of income [1]. For a prognosis of rail transport demand it is necessary to research some macro- and micro-economic indicators.

When a model of rail transport demand (expressed in rail transport by transport performances statistically) is created by statistical methods it is necessary to have access to internal and external data. Internal data are usually available from company software (for example economic software, information systems and the like) [2]. Thus processed information can be used to create a regression model very simply.

Nowadays rail companies monitor a lot of basic indicators:
- number of passengers;
- transport performances in passenger or ton kilometers;
- average transport distance by type of travel in rail passenger transport or commodities in rail freight transport;
- transport performances in train kilometers;
- transport performances in gross ton kilometers;
- fulfillment of train traffic diagram;
- change of internal and external own prices;
- expenses and revenues;
- quantity and utilization of wagons, locomotives, electric or diesel motor units etc.

Rail companies sometimes carry out marketing surveys to obtain relevant data about the trip purpose, quality of services and so on.

In the case of external information, which is needed to create a prognosis for rail transport performances, a problem may occur not only in collecting it, but in its informative value and the possibility of using it in mathematical/statistical models [3].

The most important data problems include:
If a prognosis for rail transport demand for the whole territory is created, problems can “only” be related to subject comparability [4]. In the case of the creation of a prognosis for rail transport demand, the intensity of the regression model is:

Before using the multiple linear regression model to create a prognosis for rail transport demand, the intensity of the dependence between the independent variable model parameters must be detected, i.e. a multicollinearity test has to be performed.

The main causes of multicollinearity can be described as follows:

1. in the model there are a lot of independent parameters that describe the same factors (for example, average monthly wage and net income per household member);
2. the existing dependence amongst independent parameters (for example, the development of road freight transport and GDP);
3. the character of the model (for example, using the polynomial function of higher degree or physical limitation in the data model);

Before using the multiple linear regression model to create a prognosis for rail transport demand, the intensity of the dependence between the independent variable model parameters must be detected, i.e. a multicollinearity test has to be performed.

The main causes of multicollinearity can be described as follows:

1. in the model there are a lot of independent parameters that describe the same factors (for example, average monthly wage and net income per household member);
2. the existing dependence amongst independent parameters (for example, the development of road freight transport and GDP);
3. the character of the model (for example, using the polynomial function of higher degree or physical limitation in the data model);

III. 2. MULTIPLE REGRESSIVE MODEL

Rail transport demand is influenced by several indicators simultaneously. Therefore it is appropriate to use multiple regressive model for this prognosis. The general form for this multiple linear regression model is:

Before using the multiple linear regression model to create a prognosis for rail transport demand, the intensity of the dependence between the independent variable model parameters must be detected, i.e. a multicollinearity test has to be performed.

The main causes of multicollinearity can be described as follows:

1. in the model there are a lot of independent parameters that describe the same factors (for example, average monthly wage and net income per household member);
2. the existing dependence amongst independent parameters (for example, the development of road freight transport and GDP);
3. the character of the model (for example, using the polynomial function of higher degree or physical limitation in the data model);
If a group of independent parameters are not correlated, scattering parameter RFV is equal to 1. The model can be used to create a prognosis if RFV is lower than 5. If the RFV is higher than 5, the regression model is characterized by multicollinearity [6].

The problem of multicollinearity can be solved in different ways as follows [7]:

- the omission of one of the independent parameters – the simplest solution. If there is strong statistical dependence between two parameters, one of them can be left out;
- the use of a priori information – information about the change of one independent parameter (b1) in response to the change of another independent parameter (b2) if the other independent parameter changes in terms of the unit. The multiple regression model then has one less independent parameter; other parameters are estimated using the LSM and the independent parameter b1 from identified a priori information. The problem is obtaining adequate a priori information; the source can be previous empirical analysis;
- the combination of cross-sectional data and time series data – additional information is obtained from other types of data;
- the transformation of independent parameters – for example, delay of the original one-period model, e.g. using the model in the shape of the first differences;
- obtaining additional data – multicollinearity problem is a problem of choice. In practice, it is often not possible to obtain additional data, especially if they are annual data;
- methods of distorted estimation parameters – LSM is replaced by another method for the estimation of parameters, for example the ridge regression method, the principal components method, the method of general inversion and so on.

IV. DEVELOPMENT OF FACTORS THAT INFLUENCE RAIL TRANSPORT DEMAND

The main factors that influence rail transport demand are [8]:

- GDP;
- average monthly wage;
- ownership of a car;
- price of rail transport services etc.

The development of these factors from 2000 to 2015 is shown in the following table.

<table>
<thead>
<tr>
<th>Year</th>
<th>GDP per inhabitant</th>
<th>average monthly wage</th>
<th>number of passenger cars</th>
<th>average price</th>
<th>traffic performance of passenger car</th>
<th>unemployment rate</th>
<th>traffic performance of road public transport</th>
<th>traffic performance of rail passenger transport</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>€/inhabitant</td>
<td>€</td>
<td>number</td>
<td>€/pas. km</td>
<td>million passenger km</td>
<td>%</td>
<td>million passenger km</td>
<td>million passenger km</td>
</tr>
<tr>
<td>2000</td>
<td>5 269.0</td>
<td>33.9</td>
<td>1 274 244</td>
<td>0.0351</td>
<td>23 929</td>
<td>17.9</td>
<td>8 435</td>
<td>2 870</td>
</tr>
<tr>
<td>2001</td>
<td>6 070.0</td>
<td>380.5</td>
<td>1 292 843</td>
<td>0.0351</td>
<td>24 056</td>
<td>19.2</td>
<td>8 253</td>
<td>2 805</td>
</tr>
<tr>
<td>2002</td>
<td>6 670.0</td>
<td>433.7</td>
<td>1 326 891</td>
<td>0.0351</td>
<td>24 978</td>
<td>18.5</td>
<td>8 236</td>
<td>2 682</td>
</tr>
<tr>
<td>2003</td>
<td>7 310.0</td>
<td>436.3</td>
<td>1 356 185</td>
<td>0.0440</td>
<td>25 224</td>
<td>17.4</td>
<td>7 757</td>
<td>2 316</td>
</tr>
<tr>
<td>2004</td>
<td>8 100.0</td>
<td>485.9</td>
<td>1 197 030</td>
<td>0.0439</td>
<td>25 332</td>
<td>18.1</td>
<td>7 882</td>
<td>2 228</td>
</tr>
<tr>
<td>2005</td>
<td>9 110.0</td>
<td>557.9</td>
<td>1 303 704</td>
<td>0.0439</td>
<td>25 824</td>
<td>16.2</td>
<td>7 525</td>
<td>2 166</td>
</tr>
<tr>
<td>2006</td>
<td>10 110.0</td>
<td>622.8</td>
<td>1 333 749</td>
<td>0.0439</td>
<td>26 342</td>
<td>13.3</td>
<td>7 665</td>
<td>2 194</td>
</tr>
<tr>
<td>2007</td>
<td>11 520.0</td>
<td>668.7</td>
<td>1 433 926</td>
<td>0.0439</td>
<td>25 994</td>
<td>11.0</td>
<td>7 596</td>
<td>2 148</td>
</tr>
<tr>
<td>2008</td>
<td>12 290.0</td>
<td>723.0</td>
<td>1 544 888</td>
<td>0.0452</td>
<td>26 395</td>
<td>9.6</td>
<td>6 446</td>
<td>2 279</td>
</tr>
<tr>
<td>2009</td>
<td>11 920.0</td>
<td>744.5</td>
<td>1 589 044</td>
<td>0.0463</td>
<td>26 538</td>
<td>12.1</td>
<td>4 538</td>
<td>2 247</td>
</tr>
<tr>
<td>2010</td>
<td>12 350.0</td>
<td>769.0</td>
<td>1 669 065</td>
<td>0.0463</td>
<td>26 876</td>
<td>14.4</td>
<td>4 436</td>
<td>2 309</td>
</tr>
<tr>
<td>2011</td>
<td>12 840.0</td>
<td>786.0</td>
<td>1 749 271</td>
<td>0.0463</td>
<td>26 887</td>
<td>13.2</td>
<td>4 611</td>
<td>2 431</td>
</tr>
<tr>
<td>2012</td>
<td>13 230.0</td>
<td>805.0</td>
<td>1 824 190</td>
<td>0.0536</td>
<td>26 935</td>
<td>13.6</td>
<td>4 584</td>
<td>2 459</td>
</tr>
<tr>
<td>2013</td>
<td>13 570.0</td>
<td>824.0</td>
<td>1 879 759</td>
<td>0.0536</td>
<td>27 155</td>
<td>14.1</td>
<td>4 338</td>
<td>2 485</td>
</tr>
<tr>
<td>2014</td>
<td>13 970.0</td>
<td>858.0</td>
<td>1 949 055</td>
<td>0.0536</td>
<td>27 251</td>
<td>12.8</td>
<td>4 495</td>
<td>2 583</td>
</tr>
<tr>
<td>2015</td>
<td>14 440.0</td>
<td>883.0</td>
<td>2 027 023</td>
<td>0.0536</td>
<td>27 301</td>
<td>11.5</td>
<td>4 590</td>
<td>3 450</td>
</tr>
</tbody>
</table>

* Source: Processed data from the Statistical Yearbook and Transport Authority [9], [10], [11]
We researched factors that are available in material and documents of the Statistical Office of the Slovak Republic or company material available on their web pages.

V. RESULTS AND DISCUSSION

The research into multicollinearity significance was carried out by calculation of the scattering parameter RFV_k for the data shown in Table 2. Scattering parameter RFV_k was calculated by formula 4.

**TABLE II. THE RESEARCH INTO MULTICOLLINEARITY SIGNIFICANCE**

<table>
<thead>
<tr>
<th>parameters correlation</th>
<th>RFV_k</th>
<th>parameters correlation</th>
<th>RFV_k</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1x2</td>
<td>0.99573</td>
<td>234.42991</td>
<td>x3x7</td>
</tr>
<tr>
<td>x1x3</td>
<td>0.89193</td>
<td>9.25338</td>
<td>x3x6</td>
</tr>
<tr>
<td>x1x4</td>
<td>0.88155</td>
<td>8.44258</td>
<td>x3x7</td>
</tr>
<tr>
<td>x1x5</td>
<td>0.96155</td>
<td>26.00797</td>
<td>x3x5</td>
</tr>
<tr>
<td>x1x6</td>
<td>-0.83154</td>
<td>5.93599</td>
<td>x3x6</td>
</tr>
<tr>
<td>x1x7</td>
<td>-0.88420</td>
<td>8.63528</td>
<td>x3x7</td>
</tr>
<tr>
<td>x2x3</td>
<td>0.90841</td>
<td>10.91862</td>
<td>x4x7</td>
</tr>
<tr>
<td>x2x4</td>
<td>0.87436</td>
<td>7.95953</td>
<td>x4x5</td>
</tr>
<tr>
<td>x2x5</td>
<td>0.96896</td>
<td>32.22114</td>
<td>x4x7</td>
</tr>
<tr>
<td>x2x6</td>
<td>-0.78413</td>
<td>4.63243</td>
<td>x5x7</td>
</tr>
<tr>
<td>x2x7</td>
<td>-0.91094</td>
<td>11.22826</td>
<td>x5x7</td>
</tr>
</tbody>
</table>

Source: Authors

As can be seen in Table 2, the coefficient RFV_k has a high level in many pair combinations. It is caused by the dependence of macroeconomic indicators on each other (for example, in the Slovak Republic there is strong dependence between GDP and average monthly earnings.

Scattering parameter RFV_k is high in most of the studied parameters. When we increased the number of parameters (data have been investigated since 1995) the dependence among parameters was higher. It follows that the regression model can be created (in our study factors) with three independent variants (every variant has three independent parameters), namely:

1. GDP per inhabitant, number of passenger cars, average price of rail passenger transport;
2. number of passenger cars, average price of rail passenger transport, unemployment rate, and
3. average price of rail passenger transport, unemployment rate, traffic performance of road public transport.

Because the traffic performance of road public transport and the traffic performance of rail passenger transport have decreased and the reason for this decrease is an increase in the standard of living, the best model for creating a prognosis for rail transport demand is variant 2 or 3. This selection is usable only in the researched conditions. If the conditions change, new research will need to be carried out.

VI. CONCLUSION

Mathematical/statistical models are one of the tools that facilitate manager decisions. However, it is not always possible to establish a model that would include all the factors that affect the decision-making process, therefore the creation of the model would be very difficult. One of the solutions is to create multiple models with partial indicators. Even if a model were created with high validity, it should not be forgotten that the passengers decide on the choice of transport means. Their preferences may be different in the future than they were in the past. It is very important to know the reason for choosing transport means. Nowadays the creation of a reliable prognosis requires a combination of mathematical/statistical and expert methods.

ACKNOWLEDGMENT

This paper was created within the framework of the following projects:

Project VEGA 1/0701/14 “The impact of the railway freight transport market liberalization on the social transport costs.”

Project VEGA 1/0095/16 “Assessment of the quality of connections on the transport network as a tool to enhance the competitiveness of public passenger transport system”.

REFERENCES